图书情报工作
第63卷 第19期 2019年10月

基于ERGM的学科交叉领域知识连接机制实证研究*

操玉杰[1] 李纲[1] 毛进[1] 杨冠灿[2]

[1] 武汉大学信息资源研究中心 武汉 430072 [2] 中国人民大学信息资源管理学院 北京 100872

摘要：［目的/意义］旨在通过探讨学科交叉领域共词网络生成的影响因素及其作用机理，揭示学科交叉领域的微观知识连接机制。［方法/过程］结合网络嵌入性理论，将学科交叉领域关键词共现关系建立的影响因素归纳为网络结构因素（内生变量）和关键词属性因素（外生变量），进而借助指数随机图模型，选择学科交叉领域“医学信息学”开展实证研究。［结果/结论］研究结果表明：网络结构对共现关系生成的影响大于关键词本身属性的影响；择优连接机制和传递性机制具有显著正向作用；关键词节点倾向于与较新节点相连；医学信息学的关键词倾向于与基础学科的关键词建立共现关系，而基础学科的关键词却倾向于与自身学科关键词相连。

关键词：指数随机图模型 跨学科 同质性 择优连接 时间效应

分类号：G250.2

DOI：10.13266/j.issn.0252-3116.2019.19.013

1 引言

跨学科研究已成为科学研究活动的重要形式，学科交叉研究领域业已占据科学版图的较大比例且处于重要位置[1]。不仅在自然科学门类内部发生着学科交叉研究，近年来，社会计算、数字人文、计算传播学等跨越学科门类的交叉领域亦受到研究者的重视。学科交叉现象引起了科学学和科技哲学等学科的关注，同时在情报学领域，也产生了以科技文献分析为手段、以学科交叉领域为对象的相关研究[2-3]。通过文献计量来揭示期刊[4]、学者[5]、研究领域[6]的学科交叉特征是其中的主流研究方向。

近年来，较多研究采用科学知识网络描述科学知识系统，根据知识节点及其关系类型的不同，涉及引文网络、合著网络、共词网络等具体网络模型[7-8]。科学知识网络中的知识连接，起到了关联不同知识元素的作用，是知识网络的微观结构基础。研究科学知识网络中知识连接的形成过程和动力学机制，有助于从微观视角深入理解科学知识的产生、创新和演化规律。共词网络是一种重要的知识网络，其节点代表主题、含有语义内容的关键词[9]。从时序角度来看，领域关键词的产生存在先后关系，因而建立共现关系的关键词之间因时间先后而发生着知识连接现象。基于此，笔者采用共词网络研究学科领域中的知识连接。

学科领域的共词网络以代表学科主题的关键词为基础，从某种程度上体现了学科领域的知识结构，能够揭示其内部知识关系[10]。目前，较多研究利用共词网络的语义信息揭示领域的主题结构及其演化[11-13]。学科交叉领域是由多个传统学科跨越学科边界而发展起来的[2]。针对学科交叉领域，一些研究也关注到了其特殊性，在主题分析时考虑到主题形成过程中来自不同学科的影响，以理解跨学科知识如何在学科交叉领域发生作用[14-15]。然而，当前的主题分析仍倾向于宏观趋势研究，尚未细致到微观作用过程的分析。事实上，共词网络本身承载着微观知识的相互关系，研究共词网络的结构成因是解剖领域知识系统的微观作用机制的有效手段。当前，有些研究已借助复杂网络和社会网络等分析方法揭示共词网络的结构特征，例如发现了共词网络中的小世界现象、节点度的幂律分布规律等[16]。然而，现有研究主要存在两方面的不足：①这些分析多关注网络结构特征，而忽略了关键词节点本身属性的影响；②尚未针对学科交叉领域共词网

* 本文系国家自然科学基金青年项目“基于学术异质网络表示学习的知识群落发现”（项目编号：71804135）和中国博士后科学基金项目“融合语义与关系的科研社群识别与演化研究”（项目编号：2018M630885）研究成果之一。

作者简介：操玉杰（ORCID：0000-0002-8899-9626），博士研究生；李纲（ORCID：0000-0001-5573-6400），教授，博士生导师；毛进（ORCID：0000-0001-9572-6709），副研究员，博士后，通讯作者，E-mail：danveno@163.com；杨冠灿（ORCID：0000-0002-1706-1884），讲师，博士。

收稿日期：2018-12-06 修回日期：2019-04-21 本文起止页码：128-135 本文责任编辑：徐健

络的特殊性开展专门研究,较少考虑多学科特性对共词网络微观结构的影响。基于以上分析,学科交叉领域共词网络的结构特征能够反映多学科知识的相互作用关系,通过研究共词网络中知识连接的生成过程能够揭示跨学科知识的微观作用机理。

指数随机图模型(Exponential Random Graph Model, ERGM)是在社会网络统计分析模型基础上发展起来的一种以关系形成为对象的研究方法,旨在通过统计的方法量化分析关系形成的影响因素[17]。该模型不仅考虑了网络内生结构的影响,同时还分析节点本身的属性,能够较为全面地揭示网络生成的影响因素及其作用机理。ERGM常被用于解释社会网络的形成机制。例如,针对社会网络,利用ERGM研究青少年同伴网络形成过程中的互惠效应、传递机制和结构扩展机制[18]。一些学者逐渐将ERGM应用于知识网络的研究中,从知识网络结构特性、社会因素、语义因素等方面探讨合作网络、引文网络等的形成机制。C. Zhang等[19]借助该模型研究传递机制、优先链接机制两种网络机制和作者生产力、影响力、研究主题和性别等作者属性的同质性对于作者合作关系形成的影响机制;杨冠灿等[20]运用ERGM从引文网络的连边、度分布、传递闭合等网络结构以及专利的地域、领域、学科、所属机制和审查员类型等属性对专利引用关系的形成进行了实证性解释。然而,目前尚未有利用该模型研究共词网络的形成过程的研究。鉴于ERGM的优势和当前研究现状,笔者借助指数随机图模型揭示学科交叉领域的共词网络生成机制,较为全面地剖析网络结构、学科属性、时间等多个因素对学科交叉领域知识系统形成的影响。

2 研究方法

2.1 分析框架与研究假设

早期,社会网络生成机制的研究多采用线性回归的思路,其关注点在于节点对的属性对于关系生成的作用,未考虑其他网络结构的影响[21]。网络嵌入性理论则认为,网络行动者的行为和影响嵌入在网络环境之中,透过网络结构可以认知行动者的行为[22]。基于这种思想,影响学科交叉领域共词网络中关系生成的因素,不仅包括节点对之间的因素,同时还需要考虑节点嵌入在整体网络之中的结构信息。从网络系统角度来看,共词网络中关键词建立连边的影响因素主要来自于两个方面:①内生变量,即结构嵌入,主要是节点身处于网络中所具有的结构属性;②外生变量,即节点或者连边本身对网络连边产生影响的属性。学科交叉领域共词网络的生成机制分析框架如图1所示:

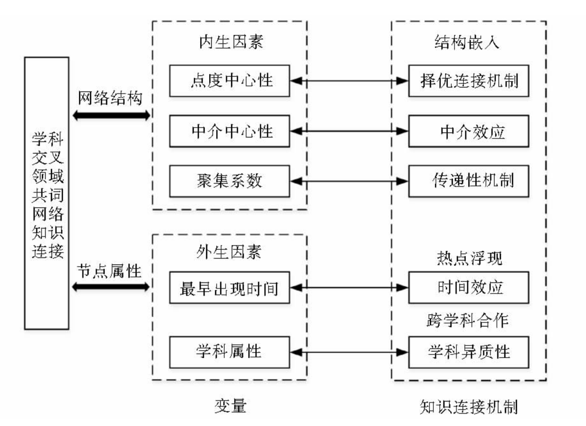


图1 理论分析框架

图1从网络内生和外生两个角度研究关键词之间如何建立共现关系。从内生角度来看,网络连边的生成受到择优连接机制、传递性机制影响。择优连接是复杂网络的一种常见动力学机制,即拥有的联系越多越容易与其他节点相连[23]。在共词网络中,关键词节点的点度中心性衡量了与其他关键词共现的次数。点度中心性越高的关键词可能代表领域研究热点,更容易产生新的知识,因此新的研究主题(可包含多个新出现的关键词)可能倾向于与这些关键词相连,从而表现出择优连接[16,24]。关键词节点的中介中心性衡量了其是否处于其他节点间的最短路径之上。处于中介位置的节点,起到了知识中介的作用,连接着不同的知识片区。在学科领域发展过程中,新的主题可能会与中介节点相连,从而使不同知识片区连接更为紧密,使得领域知识结构更加紧密。基于以上分析,笔者提出如下研究假设:

H1:关键词节点的点度中心性越大,其他节点与之产生共现关系的可能性越大。

H2:关键词节点的中介中心性越大,其他节点与之产生共现关系的可能性越大。

在网络结构中,聚集系数通过网络中三元组的数量来衡量网络中节点聚集成团的程度。三元组节点之间彼此建立关系的结构称为“三元闭包”,该结构的形成是网络演化的一种较为普遍的动力。其通俗化的解释为:朋友的朋友倾向于建立新的朋友关系,这种机制亦称为传递性机制。在共词网络中,三元闭包结构有助于知识产生聚集效应[25],具有三元闭包形成潜力的关键词之间存在内在的知识关联性,将更可能建立起知识连接。因此,笔者提出研究假设:

H3:关键词节点的聚集系数越大,其他节点与之产生共现关系的可能性越大。

从外生角度来看,关键词本身的外部属性亦可能影响到关键词之间建立共现关系。在此,主要考虑时间效应和学科异质性的影响。在学科领域的知识增长过程中,早期出现的关键词节点更有可能是学科领域的知识基础,而新兴的知识节点依托于已有知识基础而产生关联,受到学科研究的热点浮现过程影响,知识节点可能更多与产生时间与之相近的节点相连[26]。在学科交叉领域中,多学科知识相互融合发生作用,不同学科的知识可能倾向于彼此连接。为验证以上分析,笔者提出如下假设:

H4:关键词最早出现时间的差异性越小,越有助于共现关系的建立。

H5:关键词学科属性的异质性,有助于共现关系建立。

2.2 共词网络知识连接的影响因素

根据以上理论分析框架,笔者共考察表 1 中所列的 5 个影响因素,包含 3 种网络结构变量和 2 种关键词属性变量。节点的点度中心性 degree、中介中心性 betweenness 和聚集系数 cluster_coef 采用社会网络分析方法测度。关键词在领域中出现的最早年份 min_year 依据关键词所属论文进行判定。笔者采用关键词学科隶属度[27]计算方法来获取关键词的学科分类属性。基本思路是:基于投票原则,关键词隶属于某一学科的程度与该关键词在该学科文献中出现的次数成正比,与该学科的发文规模成反比。对于关键词 K,若其在学科 T 所发表文献中出现的次数为 C_T,该学科发文总量 P_T,则关键词 K 隶属于该学科的程度 S_{KT} 为:

$$S_{KT} = C_T \sum_{i=1}^{n} \frac{P_i}{P_T} \qquad \text{公式 (1)}$$

其中,N 为学科总数。计算关键词 K 在所有 N 个学科中的隶属度,然后取最大值所对应的学科作为该关键词最终归属的学科。

表 1 共词网络知识连接的影响因素

假设	变量类型	变量名	数值类型	变量解释
H1	内生变量	degree	连续型	节点的点度中心性
H2	内生变量	betweenness	连续型	节点的中介中心性
H3	内生变量	cluster_coef	连续型	节点的聚集系数
H4	外生变量	min_year	连续型	关键词在该领域中出现的最早年份
H5	外生变量	wos_category	类别型	关键词的学科分类属性,取值范围为医学信息学、医学、卫生保健、计算机科学和统计学

2.3 指数随机图模型构建

笔者利用指数随机图模型方法来研究学科交叉领域共词网络中关键词共现关系的生成机制。

ERGM 以关系数据为基础,以依赖性条件假设为条件,选择网络局部结构作为网络统计项来观察复杂网络的整体结构特征,从而获得对于网络复杂性、关联性以及随机性的整体认知[28]。ERGM 的基础公式为:

$$P_r(Y=y) = \left(\frac{1}{k}\right) exp\{\textstyle\sum_A \eta_A g_A(y)\}$$

公式(2)

其中,求和是包含所有的统计变量 A 的加总,η_A 是对应的统计变量 A 的参数,$g_A(y) = \prod_{y_{ij}\in A} y_{ij}$ 是对应统计变量的网络统计量,k 是标准化常数,确保公式(2)为适当的 0 到 1 的概率分布[29]。在具体研究中,可以根据研究内容对统计变量集合 A 进行设计,以适应研究需要。本文模型变量集合参见表 1。

简单来说,ERGM 的核心任务就是给具有某些特定机制组合的网络赋予权值的过程。因此,公式(1)也可以写成一种条件 Logit 的形式[28, 30]:

$$Logit\left[\frac{P_r(Y_{ij}=1 \mid Y_{ij}^{\ C})}{P_r(Y_{ij}=0 \mid Y_{ij}^{\ C})}\right] = \textstyle\sum_{A(y_{ij})} \eta_A d_A(y) = \sum_{A(y_{ij})} \eta_A [g_A(Y_{ij}^+, X) - g_A(Y_{ij}^-, X)]$$

公式(3)

其中,Y_{ij} 指共词网络中一条新的共现关系出现的概率,$Y_{ij}^{\ C}$ 表示网络中除 Y_{ij} 之外的其它共现关系,Logit $\left[\frac{P_r(Y_{ij}=1 \mid Y_{ij}^{\ C})}{P_r(Y_{ij}=0 \mid Y_{ij}^{\ C})}\right]$ 计算的是一条新的共现关系出现的概率与它不出现的概率的对数比值;$d_A(y)$ 是网络统计量的变化值;η_A 是相应的估计参数。通过公式(3),可以得到在一条共现关系生成概率从 0 到 1 变化的过程中,由共词网络中相关统计变量(包括节点属性和网络结构属性)变化所引起的关系生成与关系不生成的对数比值。在指数随机图模型中,该系数也被称为对数几率。当一条共现关系的对数几率被计算出为 β,可以理解为,在关系生成概率在 0 到 1 的范围内,受特定共现网络统计变量一个单位的变化影响,该共现关系生成的概率是不生成的概率的 e^{β} 倍。

基于以上设计,借助 *R* 语言的 *statnet* 包[28]构建 *ERGM* 并求解。在实证过程中,分别研究外生变量和内生变量对网络生成的作用,构建零模型、网络结构模型、节点属性模型和综合模型 4 个实证模型进行对比研究,模型细节参见表 2。零模型仅作为参照模型,考察网络连边数的影响。网络结构模型在零模型基础上,加入内生变量。节点属性模型在零模型基础上,考

察外生变量的影响,针对学科属性既研究其一元属性,研究其二元节点间的同质性。综合模型在零模型上,同时加入内生变量和外生变量。

表2　4个实证ERGM的具体细节

模型	零模型	网络结构模型	节点属性模型	综合模型
详细公式	keyword_network ~ edges	keyword_network ~ edges + nodecov("degree") + nodecov("betweenness") + nodecov("cluster_coef")	keyword_network ~edges +nodefactor("wos_category", base = 4) + nodematch("wos_category", diff = T) +absdiffcat("min_year")	keyword_network ~edges +nodefactor("wos_category", base = 4) + nodematch("wos_category", diff = T) +absdiffcat("min_year") +nodecov("degree") + nodecov("betweenness") +nodecov("cluster_coef")
考察因素	网络结构中的边数	1. 网络结构中的边数 2. 节点在网络结构中的点度中心性 3. 节点在网络结构中的中介中心性 4. 节点在网络结构中的聚集系数	1. 网络结构中的边数 2. 节点自身的最早出现年份 －差值效应 3. 节点自身的学科分类属性 －主效应 －同质性	1. 网络结构中的边数 2. 节点在网络结构中的点度中心性 3. 节点在网络结构中的中介中心性 4. 节点在网络结构中的聚集系数 5. 节点自身的最早出现年份 －差值效应 6. 节点自身的学科分类属性 －主效应 －同质性

3　实证分析

3.1　学科交叉领域选择与数据获取

选择医学信息学(Medical Informatics)作为学科交叉领域的案例开展实证研究。医学信息学具有明显的学科交叉特性[31],其产生和发展已历时较长,具有一定的文献规模。在确定学科领域的范围时,借助布拉德福定律原理,即一个学科的绝大部分论文来自于少量核心期刊[32],采用核心期刊来定义学科范围。以*Web of Science*(WOS)数据库作为文献数据来源,针对医学信息学,确定2016年版WOS的医学信息学分类中的24个期刊作为来源期刊,共检索得到1900年至2016年的37 650条Article类型论文题录,其元数据包含论文标题、作者关键词、系统关键词、期刊、发表时间、学科分类等项。

为计算关键词的学科属性,首先需要确定医学信息学的关键关联基础学科。通过统计医学信息学的参考文献所属期刊进行确定,并参考期刊的WOS学科分类信息,共选定了医学(Medicine)、卫生保健(Health Care)、计算机科学(Computer Science)和统计学(Statistics)4个学科作为关联基础学科。分别收集4个学科的期刊论文题录信息,论文发表时间亦设为1900年至2016年。

在预处理时,识别5个学科领域的关键词,计算关键词的学科隶属度。由于作者关键词的大量缺失,笔者也使用系统关键词,并从标题中抽取名词短语作为标题关键词,集成3类关键词提升关键词覆盖范围。经标点符号替换、基于最短编辑距离的同义词发现等数据清洗工作,得到最终的关键词集合。表3列出了5个学科的期刊数、论文数、关键词数、关键词总频次等信息。

表3　5个学科的基本信息

学科类别	期刊数	论文数	关键词数	关键词总频次
医学信息学	24	37 624	126 552	407 346
医学	25	119 475	227 322	841 296
卫生保健	33	76 291	183 372	783 276
计算机科学	24	70 256	227 269	693 569
统计学	25	57 529	143 829	472 414

3.2　共词网络构建

由表3可知,医学信息学领域的关键词数量较多。一方面statnet所支持的网络规模不宜过大,另一方面关键词出现频次过低,其统计值更可能受随机因素影响。因此,有必要对关键词进行筛选。借鉴J. C. Donohuc提出的高频词低频词界分公式[33],选取前1 000个高频词。由于第995个词到第1 013个词同样拥有46的频次,截取前1 013个高频词,用于构建共词网络。如果两个关键词同时出现在一篇文章中,则认为这两个关键词之间存在一条共现关系,本研究暂不考虑关键词之间的共现强度。

医学信息学共词网络的节点数为1 013,边数140 186,网络密度0.273,网络聚集系数为0.138。由此发现,相较于一般社会网络,由高频词所构成的共词网络较为稠密。进而,分别统计1 013个关键词的出现频次、最早出现年份,并进行学科属性判定。在1 013个关键词中,关键词词频最小为46次,最大为2 616

次,平均一个关键词出现 161 次。关键词最早于 1964 年出现,最晚于 2013 年出现,时间差值则为 0－49 年。根据关键词的时间分布(见图 2),大多数关键词出现在 1984－1998 年之间,表明该领域在此阶段引入或产生的知识较多,而在 1999 年以后新知识出现较少,领域进入成熟阶段。图 3 呈现了这 1 013 个关键词的学科分布结果,医学信息学的关键词数量最多(424 个),卫生保健次之(269 个),而医学最少(83 个)。

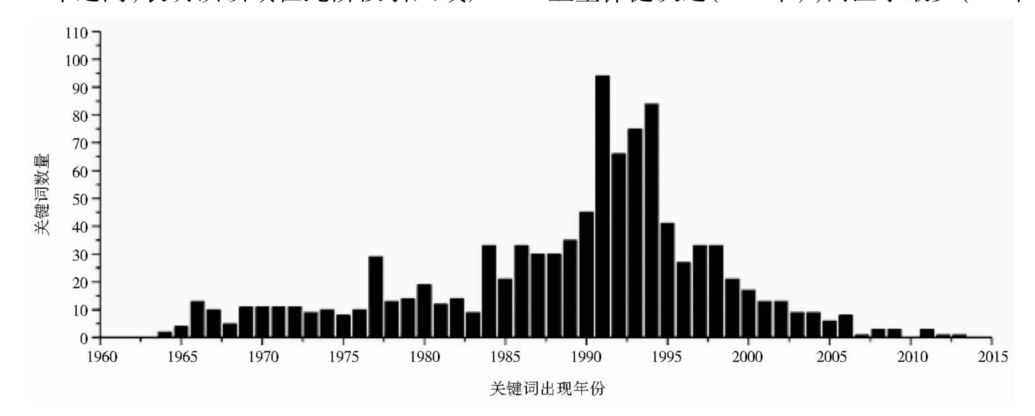


图 2 关键词最早出现年份分布

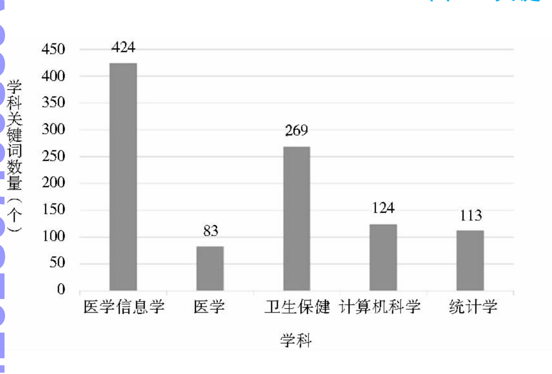


图 3 关键词的学科分布

3.3 模型结果分析与讨论

采用 statnet 工具包对零模型、网络结构模型、节点属性模型和综合模型进行参数估计,所用估计方法为马尔可夫链蒙特卡罗最大似然法(即 MCMC MLE)。模型拟合效果评价指标采用 AIC 和 BIC,指标值越小表明模型效果越好。表 4 列出了 4 个模型的参数估计结果。根据 AIC 和 BIC 值可以得出,网络结构模型的 AIC 和 BIC 值小于节点属性模型的值,表明内生因素对于关系形成的影响较之于外生因素的影响更大。从系统论角度来看,将学科交叉领域的知识系统描述为共词网络,其网络的结构特性成为关系生成的最为主要的动力因素。这从侧面反映了利用共词网络描述领域知识系统的优势。

综合模型结果最优,说明内生因素和外生因素不是独立作用于关系形成,而是相互影响的。该模型拟合数据与真实网络情况最为接近,下文主要依据综合模型的参数结果分析各个影响因素的具体作用。

3.3.1 内生变量的影响机制 在综合模型中,关键词点度中心性的拟合结果为显著正相关。该结果表明,在节点其他因素数值相同的情况下,关键词点度中心性越大,则越有可能与其他的关键词形成共现关系。这一结果印证了复杂网络中具有普适性的择优连接机制在共词网络生成中同样发挥显著作用[34],说明假设 H1 成立。这一结论与共词网络节点度数的幂律分布[16]所得结论一致。然而,关键词中介中心性的拟合结果显示为显著不相关(0.00,p-value $<$ 0.001),表明在共词网络中关键词是否处于中介位置并不影响它是否被其他关键词所链接,说明假设 H2 不成立。聚集系数的拟合结果显示为显著正相关(0.52,p-value $<$ 0.001),且该值较大。这说明,在其他因素相同的情况下,拥有较高聚集系数的关键词与其他关键词形成共现关系的可能性更高。聚集系数越大,该节点与其他节点凝聚成团的能力越大,因而新的节点与之建立连边的可能性越高,证实了假设 H3。

3.3.2 外生变量影响机制 综合模型检验了关键词节点的最早出现年份和学科分类属性两种网络外生变量对于关键词间共现关系生成的影响。针对关键词的最早出现年份变量,通过差值分析考察关键词最早出现时间如何影响共现关系生成。关键词的最早出现年份差值范围为[1－49],鉴于篇幅有限,未在表 4 中进行详尽展示,采用图 4 进行呈现。图 4 仅列出了在综合模型中参数估计结果为显著的 44 个结果,非显著的 5 个结果说明该年份差值对关键词共现关系形成的影响并不明确。

表 4　ERGM 参数估计结果

	考察变量	零模型	网络结构模型	节点属性模型	综合模型
基准模型变量	边数	−0.98 ***	−4.82 ***	−1.16 ***	−4.56 ***
内生变量	点度中心性		0.01 ***		0.01 ***
	中介中心性		0.00 ***		0.00 ***
	聚集系数		0.51 ***		0.52 ***
外生变量	主效应 − 医学信息学			参照项	参照项
	主效应 − 医学			0.30 ***	−0.13 ***
	主效应 − 卫生保健			0.14 ***	−0.32 ***
	主效应 − 计算机科学			−0.10 ***	−0.27 ***
	主效应 − 统计学			−0.34 ***	−0.36 ***
	同质性 − 医学信息学			−0.09 ***	−0.13 ***
	同质性 − 医学			0.26 ***	0.42 ***
	同质性 − 卫生保健			0.65 ***	0.86 ***
	同质性 − 计算机科学			1.09 ***	1.49 ***
	同质性 − 统计学			1.90 ***	2.28 ***
	差异性 − 最早出现年份			略	详见图 4
	AIC	601 462	476 604	585 965	466 557
	BIC	601 473	476 649	586 622	467 248

***代表 $p < 0.001$

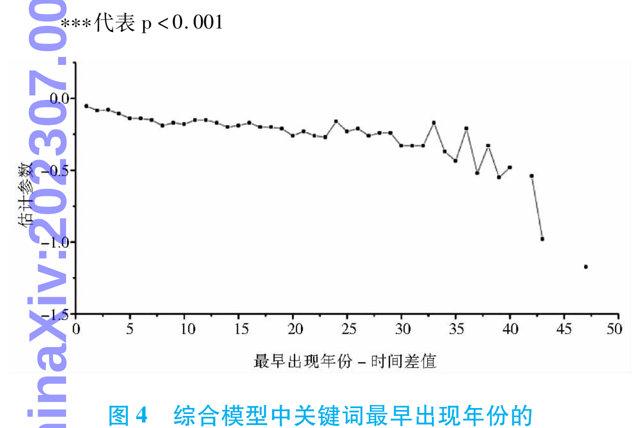


图 4　综合模型中关键词最早出现年份的时间差值效应检验结果

关键词最早出现年份的时间差值结果均为负值，说明该值的差异性对关键词形成共现关系具有显著负向影响。相对于同一年份出现的关键词，不同年份的关键词形成共现关系的可能性相对较小。同时，随着时间差值的增大，其估计参数值越来越小，表明形成共现关系的可能性也在逐渐减小。这一现象说明，在医学信息学领域，关键词倾向于与领域中出现的新关键词相连，亦表明新知识更容易发生知识连接，更可能衍生出新兴知识。该结果证实了假设 H4，可以用科学研究的研究热点浮现过程来解释：科学研究是建立在已有科学发现基础之上的不断探索的过程，新的研究发现往往会激发起更多的相关科学问题，从而形成一个个新的研究热点。这一过程在共词网络中的表现是：新的关键词倾向于与产生时间较为接近的关键词相连。

关键词节点的学科分类属性为类别型变量，共为其设计了两种检验：①主效应，考察单个节点的学科类别属性对连边生成的影响；②二元节点对之间的同质性，检验连边两端节点的学科一致性对连边生成的影响。在检验学科类别属性的主效应时，以医学信息学为参照项，拟合结果显示其他学科值均为显著负相关。该结果说明，相对于医学信息学的关键词节点，医学、卫生保健、计算机科学和统计学的关键词与其他关键词建立连边的可能性较小。反言之，在这 5 个学科中，医学信息学的关键词节点，被选择成为共现关键词的可能性最大。在整个网络中，关键词更倾向于与作为交叉领域的医学信息学关键词建立连边关系。

在表 4 中，学科分类属性的同质性检验结果是：同属于医学、卫生保健、计算机科学和统计学 4 个学科的估计参数分别为 0.86、1.49、0.42 和 2.28，均为显著正相关。由此说明，这 4 个学科的关键词更倾向于和自身学科的关键词形成共现关系。与此同时，医学信息学的同质性检验值为显著负相关（−0.13），表明同属于医学信息学的关键词建立共现关系的可能性较小，反之该学科关键词更倾向于与其他 4 个学科的关键词形成共现关系。该结论细化了假设 H5。

综上分析，医学、卫生保健、计算机科学和统计学 4 个学科的关键词，优先倾向于与本学科关键词形成共现关系，其次倾向于与医学信息学关键词形成共现

关系，表现出一定的路径倾向。然而，医学信息学的关键词，则优先倾向于与其他学科关键词形成共现关系。该说明，医学信息学这一学科交叉领域的知识倾向于与医学、卫生保健、计算机科学和统计学等其他学科知识相连，体现了该交叉领域借用、移植和融合其他学科知识的跨学科性。

4 结语

学科交叉领域的共词网络是领域内多学科知识连接的具象化呈现，共词网络中关键词共现关系是知识的微观连接的体现。笔者从学科交叉共词网络出发，运用指数随机图模型从网络内生变量和外生变量两方面入手，考察关键词共现关系建立的影响因素及其具体作用，以此理解学科交叉领域的微观知识连接机制，尤其是不同学科知识的作用。以医学信息学为例开展实证研究，得到如下发现：

(1)从网络系统角度来看，网络结构对于共现关系生成的影响大于关键词本身属性的影响。择优连接机制和传递性机制较大程度上影响了共词网络的生成。这一结果也体现了采用复杂网络描述领域知识系统的优势，借助于网络嵌入理论可以揭示领域知识系统的内在结构规律；

(2)关键词节点倾向于与较新的知识节点相连，新知识节点更容易衍生出新知识。这一规律可采用科学研究的热点浮现过程进行解释；

(3)医学信息学关键词倾向于与相关基础学科的关键词建立共现关系，反映了该交叉领域吸纳相关基础学科知识，体现了该领域的学科交叉特性。同时，还发现在这一交叉领域中，各基础学科的关键词却倾向于与自身学科关键词相连，说明交叉领域应用基础学科知识具有一定的路径性，而不是全面结合。

本文研究亦存在一些不足。仅研究了少量几个影响共词关系形成的因素，还可能存在更多的影响因素，如关键词的语义等。在实证过程中基于投票思想，采用一种学科隶属度指标进行关键词的学科分类属性判定。该方法较为简单直接，未考虑关键词属于多个学科的情况，忽略了关键词学科归属的模糊性和跨学科性现象[14, 35]。笔者仅对医学信息学这一交叉领域进行了研究，相关结论是否受学科差异性影响，尚未检验。在未来研究中，将针对这些问题进一步深入研究，并对多个交叉学科对比分析，同时将借助统计物理模型印证本文的相关发现。

参考文献：

[1] 刘仲林，赵晓春. 跨学科研究：科学原创性成果的动力之源——以百年诺贝尔生理学和医学奖获奖成果为例[J]. 科学技术与辩证法，2005，22(6)：105－109.
[2] 许海云，尹春晓，郭婷，等. 学科交叉研究综述[J]. 图书情报工作，2015，59(5)：119－127.
[3] 章成志，吴小兰. 跨学科研究综述[J]. 情报学报，2017，36(5)：523－535.
[4] LEYDESDORFF L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals[J]. Journal of the American Society for Information Science and Technology，2007，58(9)：1303－1319.
[5] PORTER A，COHEN A，DAVID ROESSNER J，et al. Measuring researcher interdisciplinarity[J]. Scientometrics，2007，72(1)：117－147.
[6] MORILLO F，BORDONS M，GÓMEZ I. An approach to interdisciplinarity through bibliometric indicators[J]. Scientometrics，2001，51(1)：203－222.
[7] 马费成，刘向. 科学知识网络的演化模型[J]. 系统工程理论与实践，2013，33(2)：437－443.
[8] YAN E，DING Y. Scholarly network similarities：how bibliographic coupling networks，citation networks，cocitation networks，topical networks，coauthorship networks，and coword networks relate to each other[J]. Journal of the American Society for Information Science and Technology，2012，63(7)：1313－1326.
[9] CALLON M，COURTIAL J P，LAVILLE F. Co-word analysis as a tool for describing the network of interactions between basic and technological research：the case of polymer chemsitry[J]. Scientometrics，1991，22(1)：155－205.
[10] 王晓光. 科学知识网络的形成与演化（Ⅰ）：共词网络方法的提出[J]. 情报学报，2009（4）：599－605.
[11] DING Y，CHOWDHURY G G，FOO S. Bibliometric cartography of information retrieval research by using co-word analysis[J]. Information processing & management，2001，37(6)：817－842.
[12] 胡吉明，张晓娟，谭婧. 我国政府信息资源研究的主题结构与演化态势[J]. 信息资源管理学报，2018，8(3)：54－63，36.
[13] 李纲，巴志超. 共词分析过程中的若干问题研究[J]. 中国图书馆学报，2017，43(4)：93－113.
[14] 许海云，郭婷，岳增慧，等. 基于 TI 指标系列的情报学学科交叉主题研究[J]. 情报学报，2015，34(10)：1067－1078.
[15] HU J，ZHANG Y. Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization[J]. Scientometrics，2017，112(1)：91－109.
[16] 王晓光. 科学知识网络的形成与演化（Ⅱ）：共词网络可视化与增长动力学[J]. 情报学报，2010（2）：314－322.
[17] ROBINS G，PATTISON P，KALISH Y，et al. An introduction to exponential random graph（p＊）models for social networks[J]. Social networks，2007，29(2)：173－191.
[18] JIAO C，WANG T，LIU J，et al. Using exponential random graph models to analyze the character of peer relationship networks and

their effects on the subjective well-being of adolescents[J]. Frontiers in psychology, 2017, 8: 583.
[19] ZHANG C, BU Y, DING Y, et al. Understanding scientific collaboration: Homophily, transitivity, and preferential attachment [J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 72 – 86.
[20] 杨冠灿, 陈亮, 张静, 等. 专利引用关系形成的解释框架：一个指数随机图模型视角[J]. 图书情报工作,2019, 63(5): 100 – 109.
[21] POL J V D. Introduction to network modeling using Exponential Random Graph Models (ERGM): theory and an application using R-Project[J/OL]. Computational Economics, 2018:1 – 31[2019 – 03 – 20]. https://doi.org/10.1007/s10614-018-9853-2.
[22] PENG T Q. Assortative mixing, preferential attachment, and triadic closure: a longitudinal study of tie-generative mechanisms in journal citation networks[J]. Journal of informetrics, 2015, 9 (2): 250 – 262.
[23] BARABÁSI A L, RAVASZ E, VICSEK T. Deterministic scale-free networks[J]. Physica A: statistical mechanics and its applications, 2001, 299(3 – 4): 559 – 564.
[24] 马费成, 刘向. 知识网络的演化 (Ⅲ): 连接机制[J]. 情报学报, 2011, 30(10): 1015 – 1021.
[25] BIANCONI G, DARST R K, IACOVACCI J, et al. Triadic closure as a basic generating mechanism of communities in complex networks[J]. Physical review E, 2014, 90(4): 042806.
[26] 马费成, 刘向. 知识网络的演化 (Ⅱ): 增长老化与知识产生时点的关系[J]. 情报学报, 2011, 30(9): 916 – 921.
[27] 吕双. 国际知识管理研究的领域分析 Ⅱ: 学科领域分布的深度挖掘[J]. 情报杂志, 2012, 31(3): 118 – 123.
[28] HANDCOCK M S, HUNTER D R, BUTTS C T, et al. statnet: software tools for the representation, visualization, analysis and simulation of network data[J]. Journal of statistical software, 2008, 24(1): 1 – 9.
[29] ROSE KIM, JI YOUN, HOWARD M, et al. Understanding network formation in strategy research: exponential random graph models[J]. Strategic management journal, 2016, 37(1):22 – 44.
[30] WASSERMAN S, PATTISON P. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p[J]. Psychometrika, 1996, 61(3): 401 – 425.
[31] 齐燕, 许海云, 方曙. 基于 WOS 数据的医学信息学学科交叉发展态势研究[J]. 中华医学图书情报杂志, 2016, 25(11): 30 – 41.
[32] 邱均平. 信息计量学[M]. 武汉:武汉大学出版社, 2007.
[33] DONOHUE J C. Understanding scientific literature: a bibliographic approach[M]. Massachusetts: The MIT Press, 1973.
[34] BARABÁSI A L. Scale-free networks: a decade and beyond[J]. Science, 2009, 325(5939): 412 – 413.
[35] KWON S. Characteristics of interdisciplinary research in author keywords appearing in Korean journals[J]. Malaysian journal of library & information science, 2018, 23(2): 77 – 93.

作者贡献说明:
操玉杰:实证分析,论文撰写;
李纲:提出论文思路,撰写论文;
毛进:提出论文思路,撰写论文,数据收集;
杨冠灿:实证分析。

An Empirical Study on Knowledge Connection Mechanism of Interdisciplinary Field Based on ERGM

Cao Yujie[1] Li Gang[1] Mao Jin[1] Yang Guancan[2]
[1] Center for Studies of Information Resources, Wuhan University, Wuhan 430072
[2] School of Information Resource Management, Renmin University of China, Beijing 100872

Abstract: [**Purpose/significance**] The article aims to explore the factors and their mechanisms influencing the generation of co-word network for interdisciplinary field, and to reveal micro-level mechanisms of knowledge connection in interdisciplinary field. [**Method/process**] Borrowing network embedding theory, the article summarizes the factors into network structure factors (endogenous variables) and keywords' attribute factors (exogenous variables). Exponential random graph model is constructed based on these factors to perform an empirical analysis on the field of Medical Informatics. [**Result/conclusion**] The results show that the influence of network structure factors on the co-occurrence relationship generation is greater than that of keywords' attributes. Preferential attachment and transitive mechanism have significant positive effect. Keywords tend to be connected with the newer ones. In addition, the keywords of Medical Informatics tend to establish co-occurrence relations with the keywords from basic disciplines, while the keywords from basic disciplines tend to be connected with the keywords in their own disciplines. The conclusions are helpful to understand the formation process of knowledge systems in interdisciplinary fields and the interactions of interdisciplinary knowledge.

Keywords: ERGM interdisciplinary homophily preferential attachment time effect